

# Submitting Nucleotide Sequences to NCBI: GenBank and Sequence Read Archive

Your data type determines your path and requirements to submit nucleotide sequences to GenBank and/or Sequence Read Archive (SRA). This document outlines paths for [Large Projects](#) (Poster 1) and [Standard GenBank Submissions](#) (Poster 2).

## Large Projects

In large projects you sequenced prokaryotic/eukaryotic genomes, meta-genomes, or transcriptomes and individual loci with more than 2500 sequences per locus. In all large-project submission paths, you must register your metadata and your biological sample(s) in the [BioProject](#) and [BioSample](#) databases. In most cases, you can start by submitting your sequence data through one of the tools within the [NCBI Submission Portal \(SP\)](#) and register BioProject/BioSample during your sequence submission process.

The chart to the right outlines submission paths for individual data types. Descriptions below follow the order in the chart, from left to right.

### Unassembled sequences (reads)

If you have only next generation sequencing (NGS) reads for your genome/metagenome, transcriptome, or targeted loci project, use the [SRA](#) tool in the [SP](#) to upload your reads. Provide your project's metadata and sample descriptions during reads submission and register your project/sample(s).

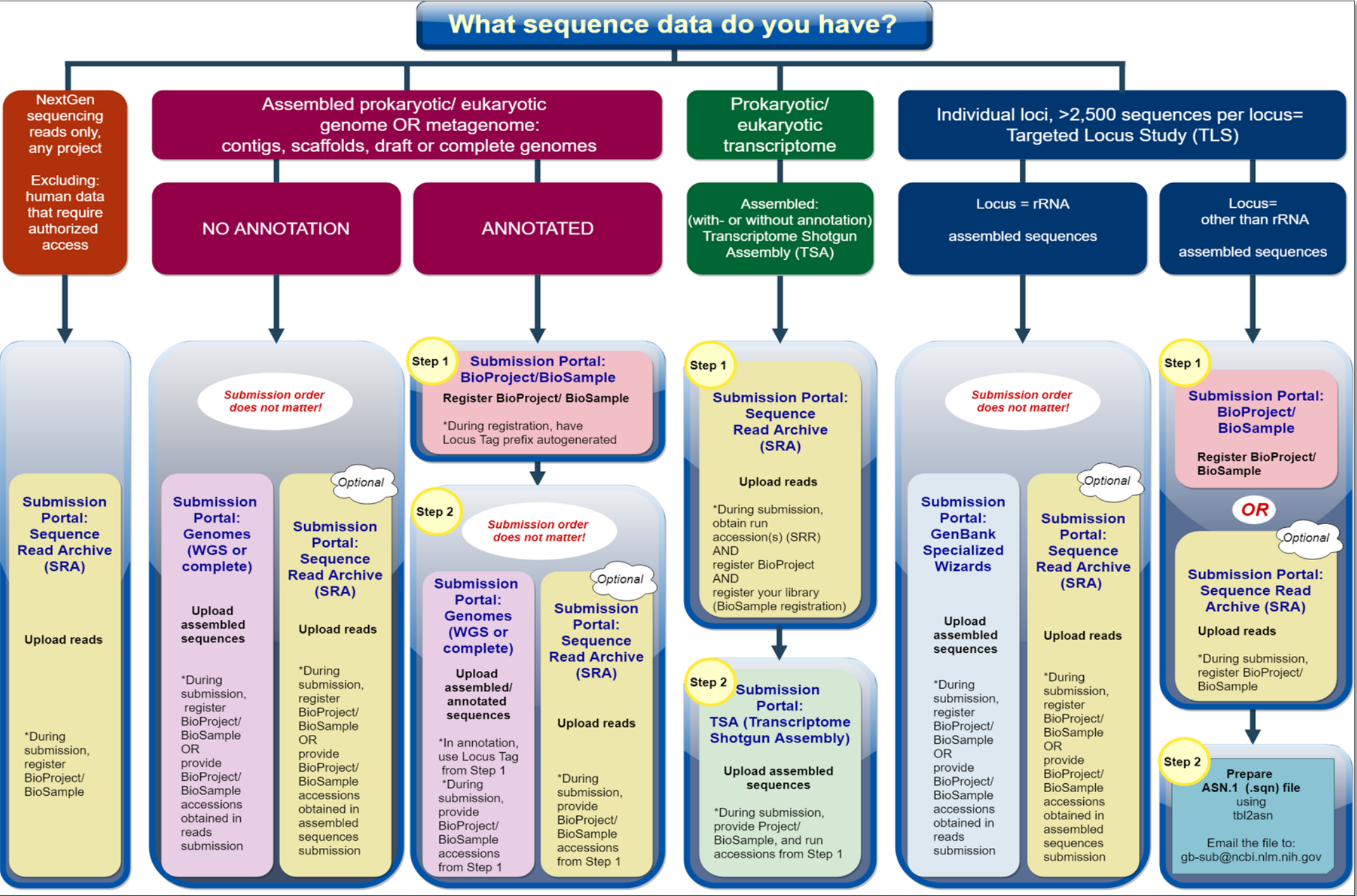
### Assembled prokaryotic/eukaryotic genomes or metagenomes — without annotation

Access the [SP](#) and select the [Genome \(WGS or complete\)](#) tool. In a single submission, you can submit a single genome or a batch of genomes from the same project. For prokaryotic genomes/metagenomes you can request NCBI to annotate your assembly for you through the [NCBI Prokaryotic Genomes Annotation Pipeline \(PGAP\)](#). Place this request using the [Genome \(WGS or complete\)](#) tool. Submit phage/plasmid/organelle genomes with a prokaryotic/eukaryotic genome. Otherwise use the [Standard GenBank Submissions](#) path for small genomes (see Poster 2). If you have associated NGS reads, upload these through the [SRA](#) tool in the [SP](#). Submission order of the reads and assembled sequences does not matter. During your first submission, provide your project's metadata and sample descriptions and obtain BioProject and BioSample(s) accessions. You will provide these accessions in your remaining submission.

### Assembled prokaryotic/eukaryotic genomes or metagenomes — with annotation

Submit your data in two steps. In [Step 1](#), register your BioProject/BioSample(s) through [BioProject/BioSample](#) tools in the [SP](#). (If you start in [BioProject](#), you will be directed to [BioSample](#) to register your biological sample(s). After that, you will be directed back to [BioProject](#).) You must select [Autogenerate locus tag prefix](#) in [BioProject](#) to obtain a prefix ahead of uploading your assembled/annotated sequences. Use locus tags with the prefix to label individual loci in your annotation file.

In [Step 2](#), submit assembled *and* annotated sequences through the [Genome \(WGS or complete\)](#) tool in the [SP](#). Submit any associated reads through the [SRA tool](#) in the [SP](#). Submission order does not matter. In both tools, provide the BioProject/BioSample(s) from [Step 1](#).



### Prokaryotic/eukaryotic transcriptome

Submit in two steps. In [Step 1](#), upload reads in the [SP](#) with the [SRA](#) tool. During submission, register your BioProject/BioSample(s). You will also receive your run accession(s). In [Step 2](#), submit your transcriptome shotgun assembly (TSA) with the [TSA](#) tool in the [SP](#). During TSA submission, provide your run accession(s) and BioProject/BioSample(s) accessions from [Step 1](#).

### Targeted Locus Studies (TLS)

If you sequenced an individual locus and obtained more than 2500 sequences per locus, then you have a Targeted Locus study (TLS).

Use the specialized wizards in the [GenBank SP](#) for any ribosomal RNA (rRNA) data. The [Standard GenBank Submissions chart](#) (Poster 2) lists applicable rRNA sequences. If you have NGS reads that accompany your assembled data, upload these through the [SRA tool](#) in the [SP](#). Submission order of reads and assembled sequences does not

matter. For the first submission, provide your project's metadata and sample descriptions and obtain BioProject and BioSample(s) accessions. You will provide these accessions in the remaining submission.

For TLS loci other than rRNA, in [Step 1](#) submit the reads in [SP: SRA](#) and register BioProject/BioSample in the process. If you do not have reads, register your metadata through [BioProject/BioSample](#) tools in the [SP](#).

In [Step 2](#), prepare your assembled data as an ASN.1 (.sqn) file, with the [tbl2asn](#) software. Include the BioProject/BioSample(s) accessions in your FASTA file that you will process with [tbl2asn](#). Email the ASN.1 submission file to [gb-sub@ncbi.nlm.nih.gov](#).





# Standard GenBank Submissions

Submit sequencing data as standard GenBank submissions if they do not qualify as a large project (Poster 1).

The chart to the right lists many examples of applicable data. Note that small genomes (plasmid, virus, and organelle) also qualify as standard GenBank submissions.

## BankIt

BankIt is the preferred tool/submission path for submitting standard data to GenBank. It is a web-based tool in which you will work through tabs to fulfil various submission requirements, such as providing your sequence(s) in the FASTA format. While there is no BioProject/BioSample registration for standard GenBank submissions, you will be asked to provide descriptive data for your samples, called [source modifiers](#). In BankIt, you must also provide gene and other features annotation on your sequence(s).

## GenBank Submission Portal — Specialized Wizards

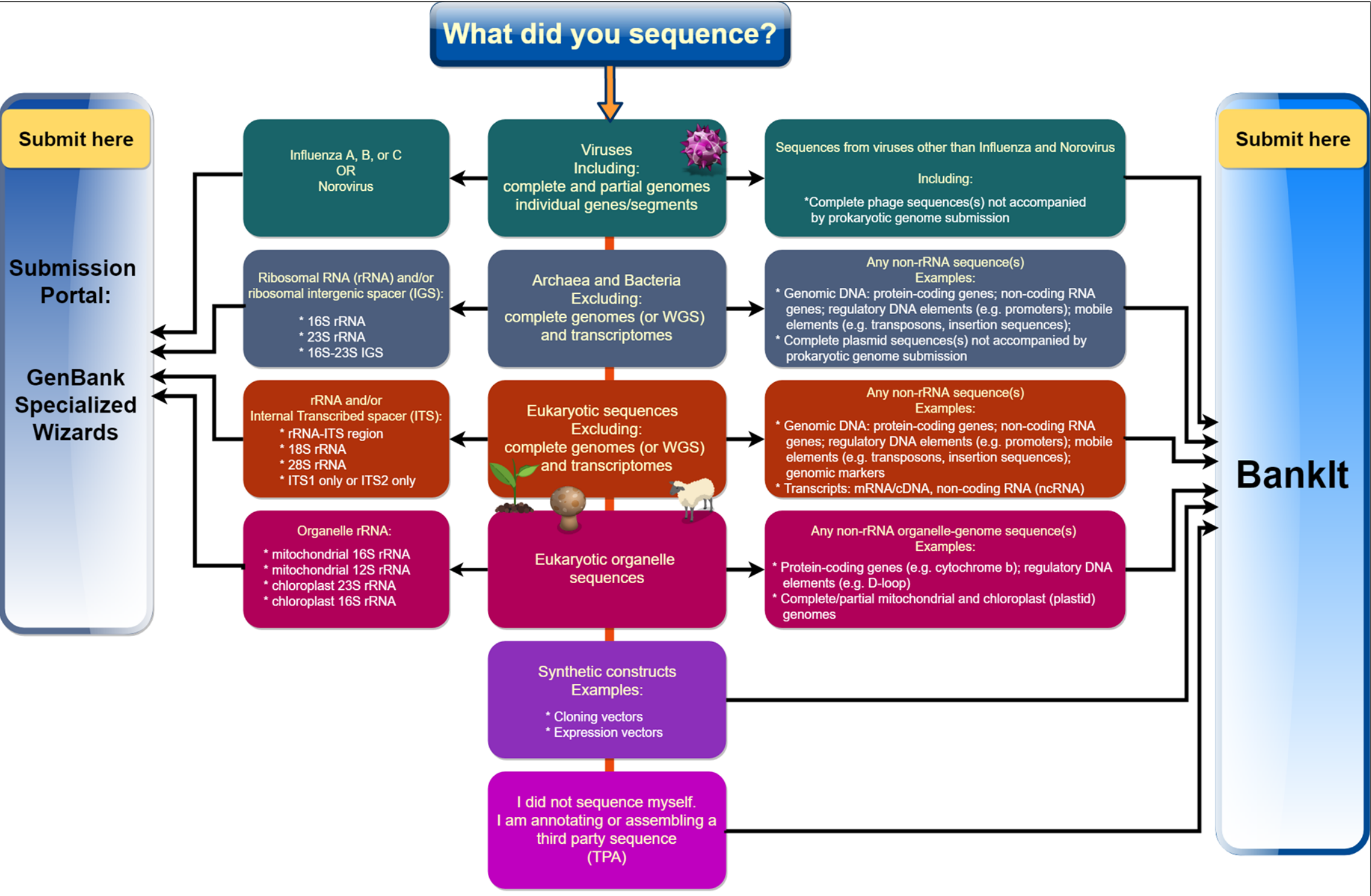
Some data types are excluded from BankIt and you must submit these through the [GenBank Submission portal, specialized wizards](#): Influenza A, B, or C, Norovirus, and ribosomal RNA (rRNA). See the chart for the listing of applicable rRNA sequences. Using the wizards will streamline your submission process as the menu options will display according to your data type. Just like in BankIt, you will import your sequence data in the FASTA format and provide any applicable source modifiers. The tool will annotate and validate your submission.

## Submitting Sets of Sequences

In a single submission (in a single BankIt or under the same “SUB” number in the GenBank Submission Portal) you can upload a set of sequences. GenBank defines a “set” as a group of sequences that represent the same gene/locus sequenced in multiple organisms. A set can also contain sequences of multiple conserved elements derived from a single organism. We publicly release all members of a set at the same time. Sets originate from population, phylogenetic, mutation, and environmental studies, that are defined as follows:

- **Population study** — same gene from different isolates of the same organism
- **Phylogenetic study** — same gene from different organisms
- **Mutation study** — multiple mutations of a single gene
- **Environmental study** — same gene from a population of unclassified or unknown organisms

If you have multiple sets, submit each set separately (one set per one BankIt number or “SUB” number in the portal). Split your submission between the GenBank Submission Portal (for rRNA) and BankIt (for other loci) if needed. For large sets (more than 2500 sequences per locus) refer to TLS (Poster 1).



### Need to update your records?

Do not prepare new submissions for records for which you already received accession numbers! Please refer to the paths/curator correspondence that you used for your original submissions to establish how you can update your accessioned records.

### Need additional help to determine your submission path/process?

Check for our contact information that we often provide within individual submission paths and guide pages. If you are unsure who to contact, write to [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov) or contact us through the NLM Support Center. We will address/direct your message as needed.

### Links to Services

**GenBank:** <https://www.ncbi.nlm.nih.gov/genbank/>

**Sequence Read Archive (SRA):** <https://www.ncbi.nlm.nih.gov/sra/>

**NCBI Submission Portal:** <https://submit.ncbi.nlm.nih.gov/>

**BankIt:** <https://www.ncbi.nlm.nih.gov/WebSub/>

**tbl2asn documentation:** <https://www.ncbi.nlm.nih.gov/genbank/tbl2asn2/>